# QUANTIFYING SOILING LOSS WITHOUT SENSORS

**João P. C. Barnabé[1], Gustavo Fraidenraich[1], João F. S. de Paula[1], João L. de S. Silva[1], Marcel V. Campos[1], Marcelo V. de Paula[1] and Tárcio A. dos S. Barros[1]**

[1] Universidade Estadual de Campinas, Campinas (Brazil)

### Abstract

Soiling is defined as the accumulation of dust, pollen, and other particulates on photovoltaic (PV) modules, and it significantly impairs energy generation by reducing light transmittance. Accurately quantifying these losses is challenging due to the dynamic interplay of environmental factors and the high cost of dedicated, permanent monitoring stations. This study proposes a novel genetic algorithm (GA) optimized approach for estimating photovoltaic soiling losses, designed to reduce reliance on physical sensors. The model predicts soiling ratios using only two widely available input variables: hourly rainfall and $PM_{10}$ particulate matter concentration. The use of a GA to optimize both the parameters and the logic of rainfall-dependent state transitions ensures high predictive accuracy without requiring complex, hard-to-measure inputs. Training was conducted on a 3-year and 5-month dataset from a 30 MW-scale PV plant, with an extra year of data being used for validation. The proposed model was benchmarked against two established models from the PVLib package: the Hsu model, used with its original parameters, and the Kimber model, optimized using the same GA for a fair comparison. The algorithm demonstrated a superior coefficient of determination ($R^2$ = 0.72), representing a performance improvement of 12% over the optimized Kimber model ($R^2$=0.59) and 32% over the Hsu model ($R^2$=0.50). By achieving high accuracy with minimal data requirements, this solution directly addresses industry priorities for low-cost, high-reliability soiling estimation.

*Keywords: Soiling, Genetic Algorithm, O&M Cost Reduction, Sensorless Monitoring, Data-driven modeling*

## 1. Introduction

Photovoltaics is the technology that generates electrical power in direct current (DC) with semiconductors by harnessing the energy from photons that illuminate them (Luque and Hegedus, 2011). As such, the performance of photovoltaic (PV) systems is degraded by soiling, which refers to the accumulation of dust, pollen, and other airborne particulates on module surfaces (Maghami et al., 2016), impeding the passage of photons through the semiconductor. In addition to degrading performance while soiling is accumulated, it also increases long term degradation (Köntges et al., 2014), may cause the PV cell to reverse bias, which possibly leads to the appearance of hotspots (Zekri et al., 2018), and recent research points to the possibility that soiling has a direct effect on module temperature itself (Sharma et al., 2024). The accurate quantification of soiling losses is essential for reliable energy yield forecasting, the optimization of operational efficiency, and the planning of cost-effective cleaning schedules. Traditional methodologies for this task rely on dedicated soiling monitoring stations; however, their high installation and maintenance costs limit widespread deployment. Data-driven estimation methods present a promising alternative, yet existing models, such as those proposed by Kimber (Kimber et al., 2006), Hsu (Hsu and Boyle, 2019), and Toth (Toth et al., 2020), often exhibit limited generalizability across diverse climatic conditions and system configurations.

The model introduced by Kimber (Kimber et al., 2006) employs a data-based approach to derive the optimal parameters for a simplified soiling model. In this framework, rainfall exceeding a specific threshold is assumed to clean the PV module completely, while sub-threshold rainfall over a consecutive number of days leads to a fixed soiling accumulation. The key parameters of this model are the rainfall threshold, the number of days required to trigger accumulation, and the accumulation rate itself. In contrast, Toth's model (Toth et al., 2020) incorporates particulate matter (PM) measurements, both coarse ($PM_{10}$) and fine ($PM_{2.5}$), by making the daily soiling accumulation value dependent on this environmental data. Hsu's model (Hsu and Boyle, 2019) adopts a fundamentally different, physics-based approach. It constructs a soiling model from


10.18086/swc.2025.03.61 Available at http://proceedings.ises.org

first principles, using variables such as dust settling velocity, aerodynamic resistance, and the quasi-laminar resistance of the PV module to estimate soiling losses by calculating the density of dust particles deposited on the surface.

The approach proposed in this study aims to add an additional layer of complexity to existing models by incorporating the possibility of different possible outcomes from rain events other than fully cleaning. It does so by extending Toth's model (Toth et al., 2020) and enhancing its structure with additional logical layers, which are then optimized using measurement data. Another core objective is to utilize only readily available, open-source data as model inputs to minimize dependency on specialized sensors. Consequently, this model uses rainfall and coarse particulate matter ($PM_{10}$) data as its primary inputs; fine particulate matter (PM2.5) data was excluded as it was not available in open datasets for the studied location. The principal enhancements involve defining different soiling outcomes for various rainfall intensities. These layers model a PM-dependent soiling increase at low rainfall levels, no change in soiling loss at a moderate level, partial cleaning at a higher level, and full cleaning above a maximum rainfall threshold.

The calibration of this enhanced model is achieved through a meta-heuristic optimization framework, specifically a Genetic Algorithm (GA). This hybrid approach, also seen in Kimber's work, albeit using linear regression and not a GA for optimization, combines the generalizability afforded by a physical representation of the soiling process with the precision of data-driven parameter fitting. The results demonstrate that the proposed model achieves an $R^2$ score of 0.72 for hourly soiling loss estimation in northern São Paulo, representing a significant improvement over the benchmark score of 0.59 obtained from existing models.

Although initial soiling measurements are required for model training, the necessary sensor can be relocated after a period of data collection, thereby reducing long-term operational costs. Furthermore, the reliance on open data means that installing a dedicated solarimetric station is not required for either the training or operational deployment of this model.

This paper is structured as follows. The Methodology section details the formulation of the novel semi-empirical soiling model and the implementation of the Genetic Algorithm optimization framework, including the chromosome encoding, fitness function, and genetic operators. The Results and Validation section presents the optimized model parameters, a comprehensive performance comparison against established benchmarks using $R^2$ and Weighted Mean Absolute Error (WMAE) metrics, and a statistical analysis of significance via the Diebold-Mariano test (Diebold and Mariano, 1995). This section also provides a detailed discussion interpreting the results in the context of the model's physical logic and its performance limitations. Finally, the Conclusion summarizes the key findings, underscores the practical implications of the proposed sensor-free solution for the PV industry, and suggests targeted directions for future research.

## 2. Methodology

The framework merges physical soiling modeling with evolutionary optimization for parameter tuning. Following the findings by Micheli and Muller (2017) that indicate the parameters most closely correlated with soiling losses, the input variables include: hourly rainfall measurements [mm] and $PM_{10}$ (particulate matter with a diameter of, at most, 10μm) concentration [$\mu g/m^3$] in the air. Input data were sourced from public government repositories: hourly rainfall data from INMET (2025) and hourly $PM_{10}$ data from CETESB (2025), both provided as integer values. The rainfall gauge is situated approximately 8 km from the plant, whereas the nearest air quality monitoring station is located about 80 km away.

Reference soiling data comes from a 30 MW PV plant in the state of São Paulo, containing values measured from August 31st, 2019 to December 31st, 2023. The data provided is the soiling ratio of the power plant, defined as the ratio between the power generated by the soiled system and the power generated by an idealized clean system; by this metric, a value of 1 represents a fully clean power plant, while a value of 0 represents a fully soiled power plant, with no generation.

### 2.1. Algorithm

The model in (eq. 1) extends Kimber's approach (Kimber et al., 2006) with Toth-inspired accumulation (Toth et al., 2020), where: t denotes the time index, r(t) is the rainfall at time t, pm(t) is the $PM_{10}$ concentration at time t, δ, β, α, γ, and ζ are the parameters to be estimated and $S_t$ is the soiling loss at time t, with $S_0 = 0$. The soiling loss, defined by 1 minus the soiling ratio, is a better metric to directly estimate by this algorithm, given its additive nature.

$$S_t\big(r(t),\, pm(t)\big) \;=\; S_{t-1}\big(r(t-1),\, pm(t-1)\big) + \begin{cases} \delta \cdot pm(t) & \text{if } r(t) \le \beta \\ 0 & \text{if } \beta < r(t) \le \gamma \\ -\alpha \cdot \sum_{i=0}^{t-1} S_i(r(i), pm(i)) & \text{if } \gamma < r(t) \le \zeta \\ -\sum_{i=0}^{t-1} S_i(r(i), pm(i)) & \text{if } \zeta < r(t) \end{cases} \quad \text{(eq. 1)}$$

Soiling changes are herein computed hourly via rainfall-dependent states. A dry state is defined for when rainfall is below the first threshold, where soiling accumulates proportionally to $PM_{10}$, following what is shown by Toth et al. (2020). Two intermediate states are then added, one in which the rain is sufficient only to prevent additional soiling buildup and another where it is enough for partial cleaning to occur, removing a fixed fraction (α) of the accumulated soiling. Lastly, there is a heavy rain state, where it fully cleans the PV modules. Figure 1 properly illustrates the flow of this algorithm.

For the intermediate rain state with a fixed cleaning fraction α, different approaches have been tested, such as the fraction linearly increasing with rainfall. However, no improvements were noticed for these other approaches, indicating that the best choice might be the most simple.

A GA described below optimizes parameters θ within empirically validated bounds (see (eq. 2)) by maximizing the coefficient of determination ($R^2$) between predicted/observed soiling.

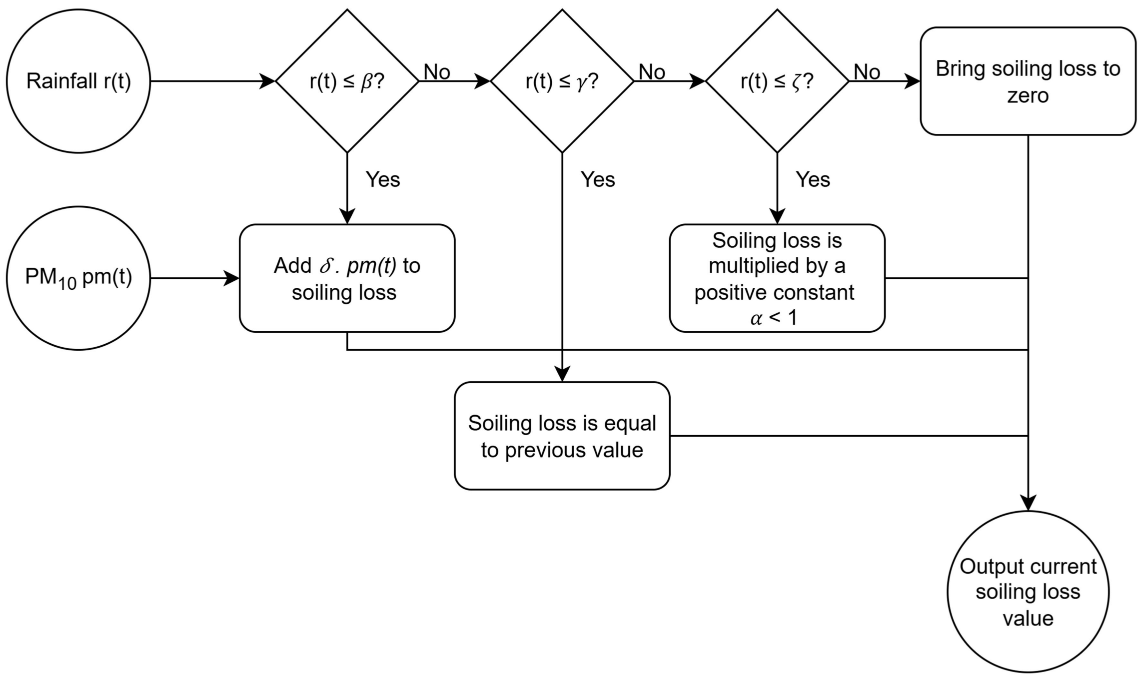


**Figure 1: Algorithm flowchart.**

$$\theta = [\delta, \beta, \alpha, \gamma, \zeta] = \begin{cases} \delta \in [10^{-6}, 10^{-2}] \\ \beta \in [0,2] \\ \alpha \in [0,1] \\ \gamma \in [0,6] \\ \zeta \in [1,20] \end{cases} \quad \text{(eq. 2)}$$

Performance was benchmarked against PVLib's (Holmgren et al., 2018) established models: Hsu (Hsu and Boyle, 2019), which used with original parameters to preserve physical fidelity; and Kimber (2006), in which the parameters were optimized identically to our method for fair comparison. An additional comparison point is a version of the algorithm that uses only rainfall data, both to test the performance in a setting with scarce data and to validate the usage of this extra parameter; the main difference in this variation is that soiling accumulation in dry hours is defined only as $\delta$ instead of $\delta \cdot pm(t)$.

## 2.2. Genetic Algorithms

Genetic Algorithms (GAs), formalized by John Holland, are a class of population-based meta-heuristic optimization algorithms inspired by the principles of natural selection and genetics (Holland, 1992). They are particularly well-suited for tackling complex, non-linear parameter estimation problems, such as tuning the inputs of physical models where gradient information is unavailable or the search space is poorly behaved.

The core metaphor consists of a population containing individuals described by a set of genes (chromosome); and their success is evaluated through a fitness function. It is applied to our problem of estimating real-valued parameters for the soiling loss model, in which an individual represents a single candidate set of parameters for the physical model; the individual's chromosome is encoded as a vector of real numbers called genes, where each gene corresponds to one parameter to be estimated. The quality of this parameter set is quantified by its fitness, defined by how well the model's output, using these parameters, matches measured empirical data via the $r^2$ score.

### 2.2.1 Implementation

At first, the initial population of size 100 is generated, with each gene of each individual initialized to a random real value within a predefined, physically plausible range for that parameter, defined as $\theta$(see (eq. 2)). Some genes in specific should be mentioned here, first, the value of $\delta$ through this algorithm is the exponent which the power of 10 is taken. Additionally, even though $\beta$, $\gamma$ and $\zeta$ define thresholds related to measured integer values, for the sake of the optimization's performance, they are kept as real values. Next, each individual's chromosome (parameter set) is passed to the physical soiling loss model. The model's output is compared against the measured soiling loss values. The fitness of the individual is calculated as the R-squared value of this comparison, quantifying the goodness of fit. Maximizing R-squared is the objective of the algorithm. As such, the individuals are ordered according to their fitness score.

Based on that score, the top 5% of individuals are automatically preserved and copied unchanged into the next generation. Guaranteeing that the best solution found is not lost from one generation to the next. From the remaining 95% of the next generation, 70% are created by selecting parents from the current population and applying genetic operators, while the bottom 25% individuals are erased and replaced by new random sets of parameters.

The first genetic operation is the crossover; in which each individual in the 70% (main parent) is replaced by an offspring produced by crossing over with a random member of the top 75%, including the fittest 5%. A uniform crossover scheme is employed: for each gene, there is a 67% probability that the offspring's gene will be calculated as the arithmetic mean of the corresponding genes from both parents and a 33% chance that the main parent's gene will remain intact.

Following that, a mutation operation is applied to the offspring, where, following crossover, each gene of the offspring has a 25% probability of undergoing mutation. A small random value, drawn from a uniform distribution with mean zero, is added to the gene. The bounds of this distribution are adaptive, getting narrower as the epochs pass. The distribution begins defined from -0.25 to 0.25, and narrows linearly until the bounds are at – 0.01 and 0.01 in the thousandth epoch.

Both crossover and mutation can act on the same gene, allowing for a robust balance between exploitation via crossover and exploration via mutation. This loop repeats for a maximum number of 1000 epochs (generations); an additional improvement based termination criterion was added, in which if the $r^2$ score of the fittest individual does not increase in 100 epochs, the process ends. A minimum amount of 250 generations was applied; in practice, all these criteria made the average number of epochs 273, running in 6 minutes, in average.

The GA optimization was also applied to the Kimber (2006) model, in which case the algorithm's hyper parameters remained the same. Each individual is encoded with only 3 parameters: cleaning threshold, which determines what level of rain should result in cleaning the PV panels; soiling loss rate, which is the rate in which the soiling loss increases in the case of there not being a cleaning event for over a determined number of days; and grace period, which is itself the number of days the algorithm waits before increasing the soiling loss.

## 2.3 Validation Strategy

To ensure a rigorous and unbiased evaluation of the proposed genetically optimized model, a structured validation strategy was employed, centered on a temporal train-test split and the use of multiple performance metrics.

### 2.3.1 Data Partitioning

The dataset, spanning from September 2019 to December 2023, was partitioned into distinct subsets for model training and final evaluation. The training set comprises all data from September 2019 to December 2022. This nearly three and a half year period was used exclusively for the calibration of the model parameters via the Genetic Algorithm. The testing set consists of the entire calendar year of 2023. This temporal split mimics a realistic operational scenario where a model is calibrated on historical data and deployed for estimating unseen data. It tests the model's performance in a situation that mimics the application proposed for it.

### 2.3.2 Performance Metrics

Model performance was quantified using two distinct metrics to provide a comprehensive view of accuracy and goodness-of-fit: the $r^2$ score and the weighted mean absolute error (WMAE). The $r^2$ value is defined as 1 minus the ratio between the mean squared error and the sample variance and can have any value between minus infinity and 1, where 1 is the result of a perfect fit; meanwhile, the WMAE is the mean absolute error calculated with weights being applied to each sample. In this work, the chosen weights are the measured values of soiling losses themselves; this was done to increase the relevance given to cases where the soiling level is higher, valuing more an estimator that can properly estimate the soiling level where it actually exists in an impactful way. Equation 3 demonstrates how WMAE is calculated.

$$WMAE = {\sum_{i=0}^{N} SM_i |SM_i - S_i|} \Big/ {\sum_{i=0}^{N} SM_i}, \qquad \text{(eq.3)}$$

where $SM_i$ is the measured soiling loss at sample i, $S_i$ is the modeled soiling loss at sample i and N is the number of samples is the validation dataset.

### 2.3.3 Benchmarking and Statistical Significance

The performance of the proposed model was compared against established benchmark models from the literature. The model of Kimber et al. (2006) was implemented and its parameters were optimized using the same Genetic Algorithm and training set (2019-2022) as the proposed model, ensuring a fair comparison of the model structures under identical optimization conditions. The physics-based model of Hsu et al. (2019) was implemented with its pre-determined parameters, representing a non-data-fitted benchmark.

To statistically validate any observed improvement in accuracy, the Diebold-Mariano (Diebold and Mariano, 1995) test was also employed, using the mean squared error (MSE) as the loss differential. This test formally assesses whether the difference in the prediction errors between two models is statistically significant. The test was conducted comparing the proposed model against the Kimber benchmark model on the 2023 test set. The final performance of all models, the proposed model and the benchmarks, was evaluated and compared solely based on their results on the unseen testing data using the two metrics described above.

# 3. Results and Validation

The evaluation demonstrates that the genetically optimized model achieves a statistically detectable and substantial improvement in estimating soiling losses compared to established benchmarks. The results confirm that the model's novel structure, which defines distinct rainfall dependent states, enables a more accurate simulation of soiling accumulation and cleaning dynamics. Furthermore, the incorporation of PM10 data is quantitatively shown to be a statistically detectable contributor to this enhanced performance.

The performance of all models was validated using hourly data from the year 2023, collected at a PV plant in western São Paulo. The final optimized parameters for the proposed algorithm are presented in Table 1, while a comparative summary of performance metrics is provided in Table 2. For a direct comparison with the measured soiling rate data, the model's soiling loss output was converted to a soiling rate value as (1 - soiling loss).

Model quality was assessed using the two previously defined metrics: $R^2$ and the Weighted Mean Absolute Error (WMAE); additionally, the Diebold-Mariano test (Diebold and Mariano, 1995) was applied to verify the statistical significance of the algorithm's results. The optimization performance of the Genetic Algorithm is illustrated in Figure 2, which shows the progression of the fitness score ($R^2$) across epochs for both the training and a separate validation dataset. It is critical to note that the validation data were held out and not used in the optimization process, demonstrating the model's ability to generalize.

An analysis of the optimized parameters in Table 1 reveals the model's learned logic for the local environment. While rainfall measurements are integers, the optimized thresholds are real-valued. This led to an unreachable state between parameters β and γ, as no integer rainfall values exist in that interval. Consequently, the system effectively operates with three distinct states: 1) For zero rainfall, soiling accumulates as a linear function of the PM10 concentration; 2) For rainfall between 1 mm and 8 mm, a partial cleaning of approximately 31% occurs; and 3) For rainfall above 8 mm, the system is fully cleaned. Furthermore, sensitivity analysis across multiple optimization runs indicated that parameters α and ζ were the most variable, suggesting the model can adapt these thresholds to different conditions. While most runs converged to the three-state system described, a minority of executions resulted in a value for γ above 1, effectively creating a fourth state; however, these configurations consistently yielded inferior performance, reinforcing the robustness of the primary three-state logic identified in Table 1.

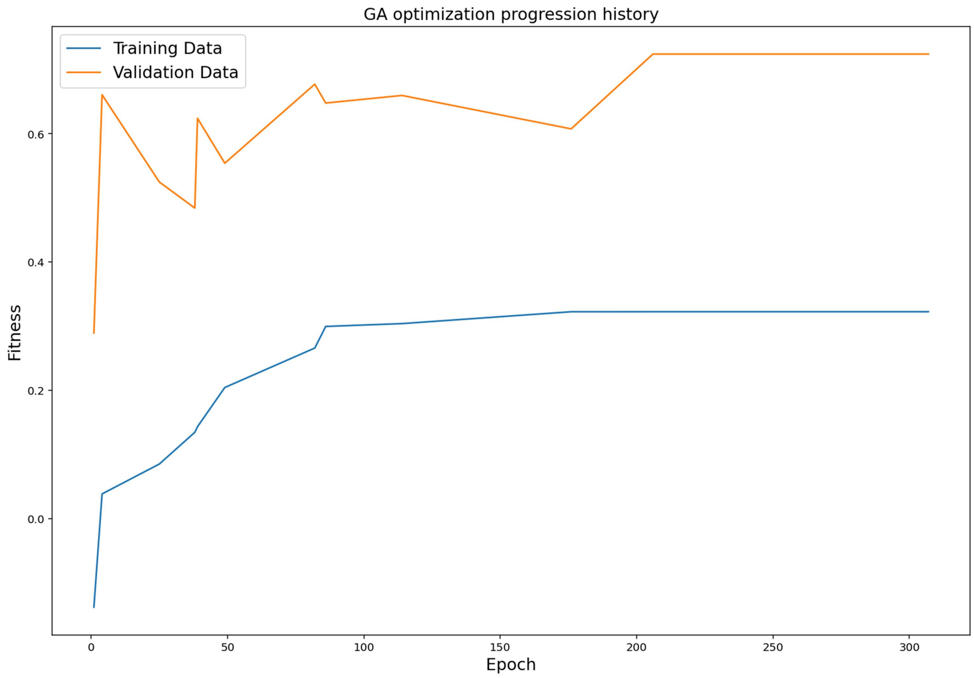


**Figure 2: Convergence of the GA optimization.**

Table 1: Optimized parameter values for the proposed algorithm.

| Parameter | Value |
|---|---|
| δ | $10^{-4.74681539}$ |
| β | 2.9 |
| α | 0.312 |
| γ | 1.4 |
| ζ | 8.11 |

Table 2: Estimation performance.

| Model | $R^2$ | WMAE |
|---|---|---|
| Proposed | 0.72 | 0.07 |
| Proposed (No $PM_{10}$) | 0.65 | 0.08 |
| Kimber et al. (2006) | 0.59 | 0.087 |
| Hsu et al. (2019) | 0.50 | 0.096 |

As summarized in Table 2, the proposed algorithm demonstrated considerably superior performance compared to the benchmark models. It achieved an $R^2$ score of 0.72, an improvement of 0.13 over the optimized Kimber model ($R^2$=0.59), and reduced the WMAE by nearly one fifth. The model without PM10 input still outperformed the Kimber benchmark, achieving an $R^2$ of 0.65, which indicates the strength of its enhanced rainfall logic alone. However, the full model's further performance gain of 0.07 confirms the significant and additive value of incorporating particulate matter data. For this specific case study, Hsu's model underperformed with an $R^2$ of 0.50. This is likely because it was developed for a different climate and geography, and its physics-based parameters were not optimized for this site. In contrast, the Kimber model presents a simpler, more generic structure that is inherently more optimizable but lacks the mechanistic detail to achieve high accuracy.

The statistical significance of the forecasting improvement achieved by the proposed model was rigorously assessed using the Diebold-Mariano (DM) test (Diebold & Mariano, 1995). The results provide overwhelming evidence for its superiority. The DM test, comparing the proposed model against the optimized Kimber model, yielded a p-value on the order of $10^{-60}$, decisively rejecting the null hypothesis of equal predictive accuracy at any conventional significance level.

Furthermore, to quantify the specific contribution of incorporating PM10 data, the proposed model was compared against a reduced version of itself that excluded the PM10 input. The DM test for this comparison resulted in a p-value on the order of $10^{-29}$, providing strong statistical evidence that the inclusion of particulate matter data enhances the model's forecasting performance. In both cases, the p-values are far below the standard threshold of 0.05, confirming that the observed improvements are not due to random chance but are statistically detectable.

Figure 3 visually underscores the improved accuracy of the proposed method by comparing the estimated and measured soiling rates over time. A key instance occurs around the 4700th hour of the year, which features a single, light rainfall event during an otherwise dry period. While the benchmark models overestimated the cleaning effect of this light rain, the proposed model accurately fit the measured data. This can be directly attributed to the optimized rainfall thresholds (Table 1), which correctly classified this event into the 'partial cleaning' regime rather than a 'full clean.' This specific example illustrates the quantitative performance gap shown in Table 2 and demonstrates the model's enhanced capability to capture complex soiling dynamics. Even more specifically, this case shows an enhancement in estimating higher values of soiling losses, most likely playing a large part in the improved WMAE value.

The superior performance of the proposed model can be directly attributed to the behavioral logic encapsulated by its optimized parameters. The benchmark models operate on simpler, more rigid assumptions, such as a single cleaning threshold. In contrast, our model's ability to define a partial cleaning state (as evidenced by the $\gamma$ and $\zeta$ parameters) allows it to correctly respond to light rain events. Benchmark models systematically misinterpret these events as full cleaning, leading to an underestimation of subsequent soiling losses, a key source of their error. This nuanced, data-driven representation of the underlying physics is the key driver behind the observed performance gains.

Despite its strong performance, this study has several limitations. The model was trained and validated on data from a single climatic region in São Paulo; its generalizability to arid or coastal environments with different soiling compositions remains to be verified. The available dataset also lacked manifestations of certain physical phenomena, such as soiling saturation (a widely reported issue where accumulation and soiling loss plateau (Kimber et al., 2006)) which consequently could not be incorporated into the current model structure. Furthermore, the model's initial dependency on a temporary, physical soiling station for training, while a reduction in long-term costs, presents a logistical hurdle for initial deployment. Finally, the use of fixed GA parameters, while simplifying the current implementation, may not be optimal for all conditions, suggesting a direction for future algorithmic refinement.

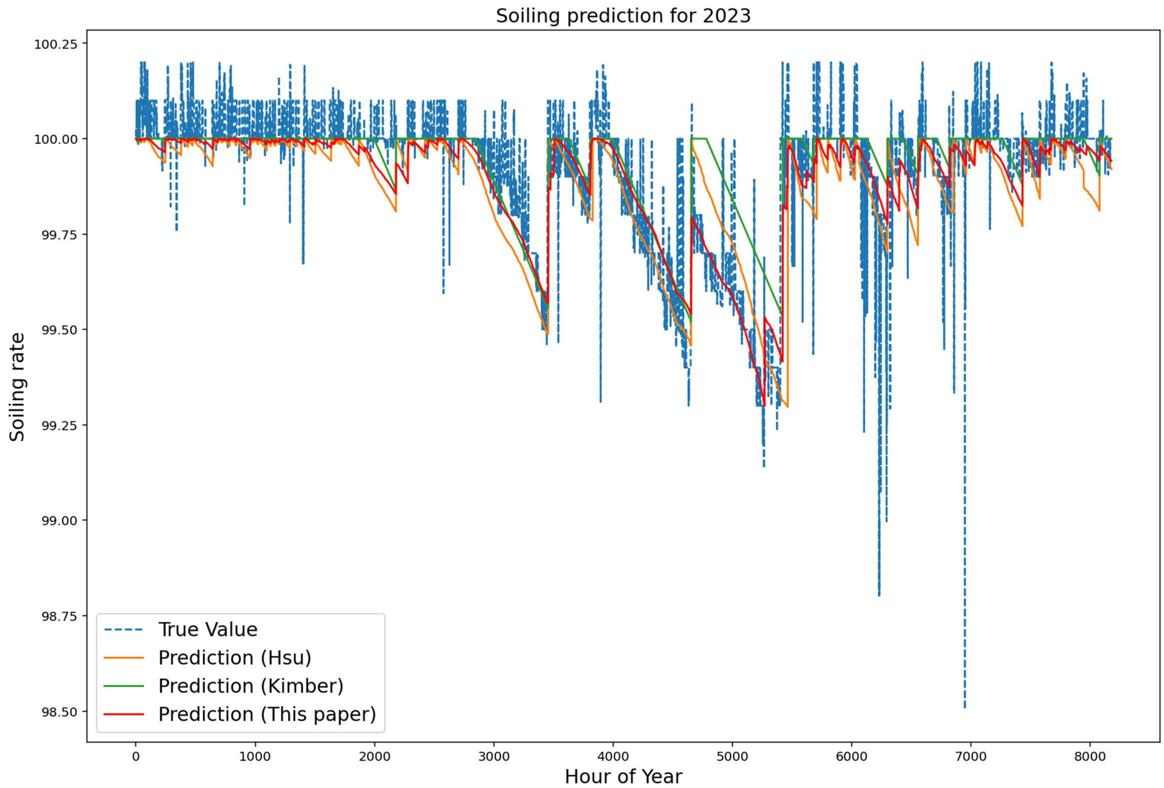


**Figure 3: Comparison of predicted vs. measured hourly soiling ratios between the three algorithms.**

## 4. Conclusion

By combining a novel, multi-state physical model with a genetic algorithm for parameter optimization, this research has developed a high-fidelity soiling estimation tool; that harness the advantages of both physics based and data driven approaches. The proposed genetically optimized, semi-empirical algorithm demonstrated a high degree of accuracy in modeling the soiling ratio, establishing a improvement over established benchmarks.

However, several important caveats must be acknowledged. First, the algorithm is designed to replicate the measurements of a specific soiling meter. Therefore, its performance is inherently limited by the accuracy and representativeness of that sensor and its installation, rather than directly quantifying the de facto soiling rate of the entire power plant. While this distinction may often be negligible, it is a critical consideration for deployment.

A second limitation, related to the first, concerns the scope of conditions captured in the training data. The model was developed and validated using data from a region characterized by relatively low soiling buildup. Consequently, physical phenomena such as soiling saturation, as predicted by models like the one in Kimber et al. (2006), were not observed in the dataset. This makes it impossible for the current algorithm to represent or validate such behavior, highlighting a dependency on the environmental conditions present in the training data.

This work naturally points to several avenues for future research. The model's architecture could be extended to incorporate a broader set of physical variables, such as relative humidity and wind speed, to achieve an even more comprehensive representation of the soiling process. It is important to note, however, that such integration is non-trivial, as preliminary attempts to include factors like tilt angle degraded performance, suggesting a need for more sophisticated functional representations. Furthermore, investigating spatial interpolation techniques to transfer learned parameters to new geographic regions without local training data represents a promising path toward true global scalability, pending the availability of comprehensive multi-location soiling datasets for validation.

Notwithstanding these limitations and future directions, the results of this work confirm that the proposed approach provides a scalable, sensor-free solution tailored to the practical needs of the PV industry. By leveraging only widely available meteorological data and a robust meta-heuristic optimization framework, this model offers a cost-effective and accurate alternative to permanent, dedicated soiling stations for ongoing soiling loss assessment, thereby facilitating the widespread adoption of precise soiling analytics and contributing to improved operational efficiency and maintenance planning for photovoltaic systems, improving profitability and grid reliability for solar assets.

## 5. References

CETESB, 2025. QUALAR: Air Quality Information System. São Paulo, Brazil. Available at: https://qualar.cetesb.sp.gov.br/qualar/home.do (accessed 20 March 2025).

Diebold, F. X., and Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253–263. https://doi.org/10.2307/1392185

INMET, 2025 – Instituto Nacional de Meteorologia. Station Map. Brasília, DF, Brazil. Available at: https://mapas.inmet.gov.br/ (accessed 20 March 2025).

Holland, J. H., 1992. Genetic Algorithms. Scientific American, vol. 267, no. 1, 1992, pp. 66–73. JSTOR, http://www.jstor.org/stable/24939139.

Holmgren, W.F., Hansen, C.W., and Mikofski, M.A., 2018. pvlib python: a Python package for modeling solar energy systems. Journal of Open Source Software, vol. 3, no. 29, page 884.

Hsu, C.-S., and Boyle, L., 2019. Simple model for predicting time series soiling of photovoltaic panels. IEEE Journal of Photovoltaics, vol. 9, no. 5, pp. 1382–1387.

Luque, A., Hegedus, S., 2011. Handbook of Photovoltaic Science and Engineering, 2nd ed., pp 1-10.

Kimber, A., Mitchell, L., Nogradi, S. and Wenger, H., 2006. The effect of soiling on large grid-connected photovoltaic systems in California and the Southwest region of the United States. In Proceedings of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion (WCPEC-4), vol. 2, pp. 2391–2395.

Köntges, M.; Kurtz, S.; Packard, C.; Jahn, U.; Berger, K. A.; Kato, K.; Friesen, T.; Liu, H.; Iseghem, M. V.; Wohlgemuth, J., 2014 Review of failures of photovoltaic modules. IEA International Energy Agency, pp 73-79.

Maghami, M. R., Hizam, H., Gomes, C., Radzi, M. A., Rezadad, M. I., Hajighorbani, S., 2016. Power loss due to soiling on solar panel: A review, Renewable and Sustainable Energy Reviews, Volume 59, pp 1307-1316.

Micheli, L. and Muller, M., 2017. An investigation of the key parameters for predicting PV soiling losses. Progress in Photovoltaics: Research and Applications. vol. 25. https://doi.org/10.1002/pip.2860.

Sharma, S., Malik, P., Sinha, S., 2024. The impact of soiling on temperature and sustainable solar PV power generation: A detailed analysis, Renewable Energy, vol. 237, Part C.

Toth, S., Hannigan, M., Vance, M. and Deceglie, M., 2020. Predicting Photovoltaic Soiling From Air Quality Measurements. In IEEE Journal of Photovoltaics, vol. 10, no. 4, pp. 1142-1147.

Zekry, A., Shaker, A., Salem, M., 2018. Chapter 1 - Solar Cells and Arrays: Principles, Analysis, and Design. Advances in Renewable Energies and Power Technologies, Elsevier, pp 3-56.